A New Method for Assessing Critical Thinking in the Classroom

A New Method for Assessing Critical Thinking in the Classroom

AHRASH N. BISSELL AND PAULA P. LEMONS

To promote higher-order thinking in college students, we undertook an effort to learn how to assess critical-thinking skills in an introductory biology course. Using Bloom's taxonomy of educational objectives to define critical thinking, we developed a process by which (a) questions are prepared with both content and critical-thinking skills in mind, and (b) grading rubrics are prepared in advance that specify how to evaluate both the content and critical-thinking aspects of an answer. Using this methodology has clarified the course goals (for us and the students), improved student metacognition, and exposed student misconceptions about course content. We describe the rationale for our process, give detailed examples of the assessment method, and elaborate on the advantages of assessing students in this manner.

Keywords: critical thinking, metacognition, Bloom's taxonomy, biology, assessment

Several years ago, we launched a journey toward understanding what it means to teach critical thinking. At that time, we were both biology instructors working together on teaching an introductory biology course at Duke University, and we wanted to help students develop higher-order thinking skills—to do something more sophisticated than recite back to us facts they had memorized from lectures or the textbook (i.e., what many of them had been asked to do in previous biology courses).

The justification for our journey is well supported by the science education literature. Many college and university faculty believe that critical thinking should be a primary objective of a college education (Yuretich 2004), and numerous national commissions have called for critical-thinking development (e.g., AAAS 1989, NAS–NRC 2003). Yet when trying to implement critical thinking as an explicit goal in introductory biology, we found ourselves without a well-defined scheme for its assessment.

And we were not alone. Despite the interest among faculty in critical thinking as a learning goal, many faculty believe that critical thinking cannot be assessed or they have no method for doing so (Beyer 1984, Cromwell 1992, Aviles 1999). Consider a 1995 study from the Commission on Teacher Credentialing in California and the Center for Critical Thinking at Sonoma State University (Paul et al. 1997). These groups initiated a study of college and university faculty throughout California to assess current teaching practices and knowledge of critical thinking. They found that although 89 percent of the faculty surveyed claimed that critical thinking is a primary objective in their courses, only 19 percent could explain what critical thinking is, and only 9 percent of these faculty were teaching critical thinking in any apparent way (Paul et al. 1997). This observation is supported by evidence from other sources more specific to the sciences, which suggest that many introductory science, technology, engineering, and math (STEM) courses do not encourage the development of critical-thinking abilities (Fox and Hackerman 2003, Handelsman et al. 2004).

Why is it that so many faculty want their students to think critically but are hard-pressed to provide evidence that they understand critical thinking or that their students have learned how to do it?

We identified two major impediments to the assimilation of pedagogical techniques that enhance critical-thinking abilities. First, there is the problem of defining "critical thinking." Different definitions of the term abound (Facione 1990, Aretz et al. 1997, Fisher and Scriven 1997). Not surprisingly, many college instructors and researchers report that this variability greatly impedes progress on all fronts (Beyer 1984, Resnick 1987). However, there is also widespread agreement

Ahrash N. Bissell (e-mail: ahrashb@duke.edu) is a research associate in biology at the Academic Resource Center, Duke University, Durham, NC 27708, where he studies teaching and learning innovation in science, as well as animal behavior and evolution. Paula P. Lemons (e-mail: plemons@duke.edu) is an assistant professor of the practice of biology in Duke's Department of Biology, where she is responsible for the required introductory biology course and works on curriculum and graduate student development.

that most of the definitions share some basic features, and that they all probably address some component of critical thinking (Potts 1994). Thus, we decided that generating a consensus definition is less important than simply choosing a definition that meets our needs and consistently applying it. We chose Bloom's taxonomy of educational objectives (Bloom 1956), which is a well-accepted explanation for different types of learning and is widely applied in the development of learning objectives for teaching and assessment (e.g., Aviles 1999).

Bloom's taxonomy delineates six categories of learning: basic knowledge, secondary comprehension, application, analysis, synthesis, and evaluation (box 1). The first two categories, basic knowledge and secondary comprehension, do not require critical-thinking skills, but the last four—application, analysis, synthesis, and evaluation—all require the higher-order thinking that characterizes critical thought. The definitions for these categories provide a smooth transition from educational theory to practice by suggesting specific assessment designs that researchers and instructors can use to evaluate student skills in any given category. Other researchers and even entire departments have investigated how to apply Bloom's taxonomy to refine questions and drive teaching strategies (e.g., Aviles 1999, Anderson and Krathwohl 2001). Nonetheless, the assessments developed as part of these efforts cannot be used to measure critical thinking independent of content.

The second major impediment to developing critical thinking in the classroom is the difficulty that faculty face in measuring critical-thinking ability per se. It is relatively straightforward to assess students' knowledge of content; however, many faculty lack the time and resources to design assessments that accurately measure critical-thinking ability (Facione 1990, Paul et al. 1997, Aviles 1999). A large body of literature already exists showing that critical thinking can be assessed (e.g., Cromwell 1992, Fisher and Scriven 1997). The critical-thinking assessments that have been most rigorously tested are subject-independent assessments. These assessments presumably have the advantage of allowing measurements of critical-thinking ability regardless of the context, thus making it possible to compare different groups of people (Aretz et al. 1997, Facione et al. 2000). Previous studies have demonstrated a positive correlation between the outcomes of these subject-independent tests and students' performance in a course or on a task (e.g., Onwuegbuzie 2001). Such studies serve to illustrate that critical thinking per se is worth assessing, or at least that it has some relationship to students' understanding of the material and to their performance on exams. Still, generalized assessments of critical-thinking ability are almost never used in a typical classroom setting (Haas and Keeley 1998). There are several problems with such general tests, including the following:

- Faculty doubt that the measurements indicate anything useful about discipline-specific knowledge.
- Administering these tests takes time away from the content of the course and can be costly; thus, they are viewed as "wasted" time.
- Most faculty lack the time to learn the underlying structure and theory behind the tests, and so it is unclear to them why such a test would be worthwhile.

Recognizing the problems with standardized, discipline-independent assessments of critical thinking, we developed an assessment methodology to enable the design of questions that clearly measure both the content we want students to know and the cognitive skills we want them to obtain. Ideally, this methodology should allow for discipline-specific (i.e., content-based) questions in which the critical-thinking component can be explicitly dissected and scored. Furthermore, we built on the work of others who have used Bloom's taxonomy to drive assessment decisions by using this taxonomy to explicitly define the skills that are required for each question. Finally, we crafted a system for developing scoring rubrics that allows for independent assessment of both the content and the skills required for each question. It is this methodology we have begun applying to introductory biology.

Box 1. Bloom's taxonomy of educational objectives.

Bloom's taxonomy subdivides the academic skills that students might need into six different categories, listed below.

1. Basic knowledge: memorizing facts, figures, and basic processes.
2. Secondary comprehension: understanding and illustrating the facts.
3. Application: generalizing the facts to other contexts and situations.
4. Analysis: understanding why the facts are the way they are; breaking problems down.
5. Synthesis: making connections between different elements on one's own.
6. Evaluation: critically using one's knowledge to ascertain the quality of information.

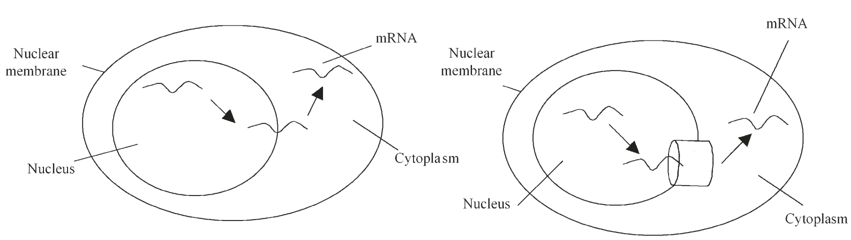
The first three categories are considered to be hierarchical: basic knowledge requires no critical-thinking skills, secondary comprehension expands on basic knowledge but also requires no critical thinking, and application requires higher-order thinking about the knowledge that a student constructs. The last three categories are also considered higher-order skills that require critical thinking, but they are not necessarily hierarchical. Note that correctly using the higher-order skills requires both knowledge and comprehension of the content, so all levels of thinking should be encouraged.

Designing, testing, and scoring discipline-specific assessments of critical thinking

Our methodology consists of several steps. First, we write questions that require both biological knowledge and critical-thinking skills. Second, we document the particular content and critical-thinking skills required (e.g., application, analysis, synthesis) and then devise a scoring rubric for the question. Our scheme is a synthesis of the work of others who have devised rubrics that independently assess either content (Porter 2002, Ebert-May et al. 2003, Middendorf and Pace 2004) or critical-thinking skills (Facione et al. 2000). Third, we subject these questions to a test of validity by submitting them for review to colleagues who are experts in biology and biological education. Fourth, we administer the assessments to students and score them on the basis of the rubric that we established in advance. On average, the first two steps of the process take about an hour for a new question; substantially less time is required when revising existing questions. For the third step, the speed at which questions can be reviewed and validated depends on the number and quality of professional colleagues, but this step is not crucial in terms of trying out a new question in a course.

Figures 1–3 illustrate three examples of our methodology. These examples are just a few of the many questions that we have already developed for use in biology. The methodology appears to be robust for different types of questions and varying degrees of open-endedness in the answers. The questions themselves (as the students would see them on an exam) are shown in figures 1a, 2a, and 3a. These questions are akin to other "advanced" questions in biology in that the students must bring several concepts together to give the correct answers. A substantial fraction of the points is based on the rationale presented in the answers, and the students are alerted to this fact ahead of time. The first step in evaluating the effectiveness of these problems is to clearly define the expectations for each question, as described in the captions for figures 1–3 (note that components of this framework borrow ideas from the work of Ebert-May and colleagues [2003]). These expectations are valid when the student gets the correct and complete answer, or when the student answers the question by drawing on the expected content. However, it is possible to apply the correct critical-thinking skills to these problems while getting some aspect of the content wrong or employing an alternative body of content. This insight is a key element of our assessment technique. Thus, we designed a grading rubric that explicitly clarifies the intersection of the content and the skills (detailed below), illustrated in figures 1b, 2b, and 3b.

a In eukaryotes mRNAs must travel through the nuclear membrane (a double lipid-bilayer) to the cytoplasm in order to be translated into proteins. Based on what you know about the chemical structure of mRNA and of lipid bilayers, choose the picture below that you think more accurately models the movement of mRNAs. Explain the rationale for your choice. (6 points)



b

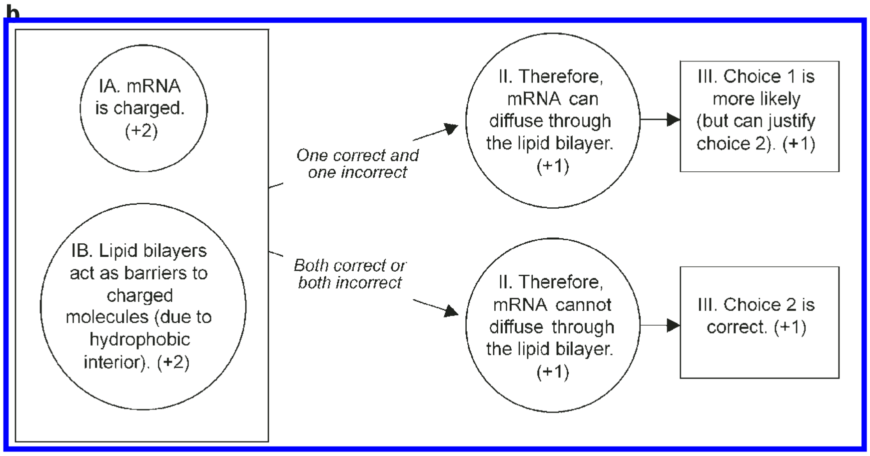


Figure 1. (a) Example question 1: mRNA diffusion through a nuclear membrane. (b) Grading rubric for example question 1. To answer this question, students need knowledge of the chemical structure of mRNA and lipid bilayers, and two types of critical-thinking skill, application *(addressing mRNA movement in the cell using knowledge of mRNA structure and lipid bilayers) and* analysis *(examining both scenarios and deciding which is more likely to be the way that mRNA moves from the nucleus to the cytoplasm). Complete answers will include each of the following elements (IA, IB, II, and III). (IA) mRNA is a molecule that has charged phosphate groups. (IB) The lipid bilayer has a hydrophobic interior. (Note: mentioning the hydrophilic head groups is not essential to this answer.) (II) The lipid bilayer cannot accomodate a charged molecule via diffusion, thereby eliminating choice 1. (III) Since diffusion won't work, an alternative mechanism is needed. Therefore, choice 2 is more likely (mRNAs move through a protein channel).*

Example 1: Chemical transport across cell membranes. A student who forgets (or never learned) about the structure of messenger RNA (mRNA) or the action of lipid bilayers can get one or both of these concepts wrong in the answer, losing two points for each incorrect part (figure 1b). However, as the rubric shows, the student can get some points if these incorrect answers are correctly rationalized to lead to an appropriate conclusion. For example, a student might indicate that mRNA is a neutral molecule (zero points) but say that lipid bilayers act as barriers to charged molecules (plus two points). In this case, the correct rationalization would be that mRNA *can* diffuse through the lipid bilayer, which means choice 1 is correct (figure 1b). The relative point values can be assigned by the professor to favor either the critical-thinking component or the knowledge needed. Thus, students can only get all of the points by knowing and understanding the content and then applying and analyzing it correctly. However, students can still get some points, even if they don't know all of the content, as long as the justification for the choice follows logically from the original errors in understanding.

Note that it is not possible to get any points for the choice if a step is missing in the rationalization. Thus, a student who correctly indicates the answers to IA and IB, and then skips the analysis, does not get any credit for the choice, even if it is correct. The reason this grading scheme works is that the choice makes no sense out of context, and the context must be provided by the student.

Example 2: Genetic drift in toads. One area of confusion that is tested by this question is the distinction between selection and drift, and the fact that environmental conditions affect these two evolutionary processes differently. Assuming the student defines genetic drift properly, then a series of logical deductions about the action of genetic drift as it pertains to the map should lead to the conclusion that the island (I) population is the correct choice. However, as figure 2b illustrates, there are many opportunities for erroneous understanding that can nonetheless be rationalized appropriately. Students may incorrectly define genetic drift (confusing it with natural selection), but then also incorrectly state that smaller populations are more likely to evolve via selection, which leads to the conclusion that species I (the island scrub toad) is most likely to change. Each part of the answer can be viewed as a decision point, and whichever decision the student makes is then carried over to the next step. In this manner, instructors can reward students who are internally consistent, even if they err somewhere early on. They also avoid rewarding students who are not consistent anywhere but happen to choose the "best" answer, which of course is only true if they get all of the components correct.

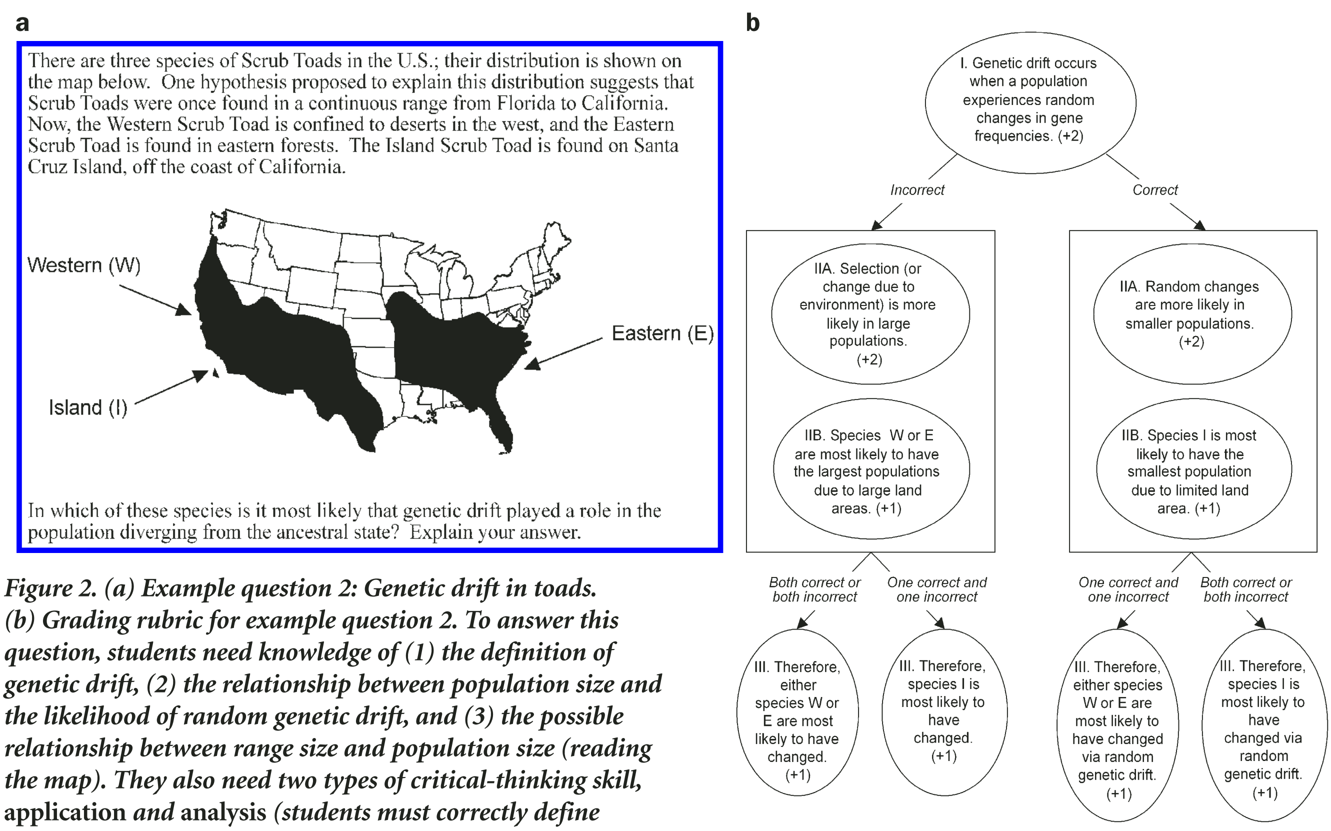


Figure 2. (a) Example question 2: Genetic drift in toads. (b) Grading rubric for example question 2. To answer this question, students need knowledge of (1) the definition of genetic drift, (2) the relationship between population size and the likelihood of random genetic drift, and (3) the possible relationship between range size and population size (reading the map). They also need two types of critical-thinking skill, **application** ***and*** **analysis** ***(students must correctly define genetic drift and state its relationship to population size, and then apply that knowledge to an analysis of likely population sizes based on the map). Complete answers will include each of the following elements (I, IIA, IIB, and III): (I) Genetic drift refers to the occurrence of random genetic changes in a population. (IIA) Random genetic changes are most likely to occur in smaller populations. (IIB) Species I has the smallest overall range, suggesting that it also has the smallest effective population size. (III) Therefore, species I is most likely to be affected by genetic drift. (Note: There are some plausible reasons why the species with larger ranges might actually have smaller effective population sizes, but the burden of fully rationalizing this conclusion falls on the student.)***

Example 3: DNA, cell division, and cancer. This question requires that students apply their knowledge of the bonding between members of DNA base pairs to understand how an anticancer drug works. It is less open-ended than the other two examples because there is really only one correct conclusion, which is essentially given within the question (i.e., nitrogen mustard is an anticancer drug). As a result, most of the errors in answering the question arise from skipping one or more of the steps necessary for describing the connection between base-pair bonding and an anticancer drug. As figure 3b indicates, the awarding of points stops when the logical flow of ideas stops; no points are awarded for the final conclusion unless it is supported by evidence presented in the answer.

As with any short-answer examination, grading is vastly easier when the students' answers are clear and employ good reasoning. But this is not typical, and the difficulty for most graders lies in the awarding of partial credit, where much time can be wasted trying to decipher whether a student's answer suggests a level of knowledge or skill that deserves some points. This methodology can greatly reduce these types of difficulties, since the range of acceptable answers should be clearly predefined. We routinely find that the questions for which we have developed a rubric require less time and effort for graders—and produce more valid scoring—than questions for which the rubric is developed during the grading process. The rubric is also subject to further refinement as a result of repeated use, in which case the time and effort needed for grading decreases even more. Of course, an instructor can save time in grading by carefully designing any type of assessment, but we believe that the level of rigor has been particularly low in assessments intended to measure critical-thinking skills, and that more explicit assessments of critical-thinking skills are needed in STEM courses.

Results

We have successfully implemented these types of assessments in a large (approximately 150 students) introductory biology course at Duke University. We are currently gathering data on student performance that allow us to assess mastery of content at several different skill levels at the same time that we test mastery of skills using different types of content. Although we have not yet completed these analyses, we have already found that the use of this assessment methodology has positively affected the course in a number of ways.

For example, thinking in advance about what we want questions to accomplish in terms of both content and critical thinking has enabled us to be explicit with students about the skills they need to develop in order to succeed in the course. We have reviewed questions and grading rubrics in our lectures and made examples of them available to students outside of class. As a result of this exposure, students were more aware of the quality of responses we expected for questions and could easily cross-reference their own responses with our explicit guidelines. These efforts helped students reflect on and improve their thinking (and writing) abilities—a concept referred to as metacognition. Conversations with students suggested that we were in fact teaching metacognition, which is known to have positive effects on learning

a

Cancer is any malignant growth or tumor caused by abnormal and uncontrolled cell division. Why does it make sense that nitrogen mustard, which causes the formation of strong covalent bonds between the members of a DNA base pair, is used as an anticancer drug?

b

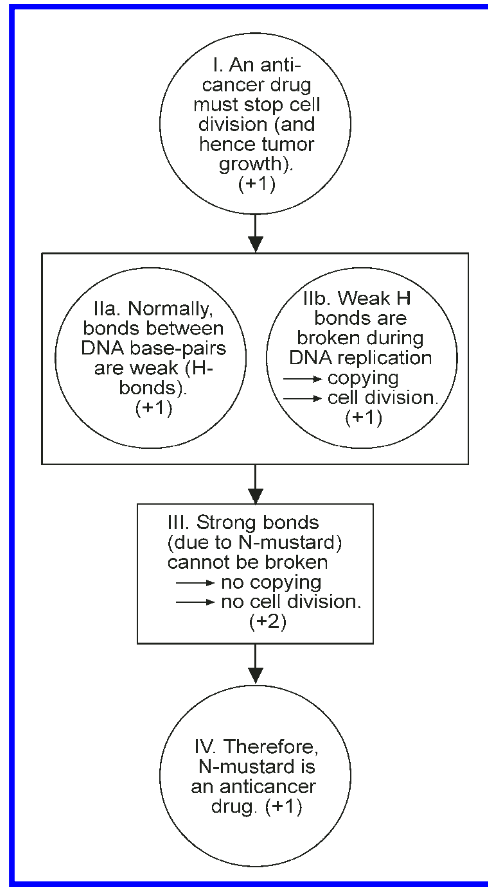


Figure 3. (a) Example question 3: DNA, cell division, and cancer. (b) Grading rubric for example question 3. To answer this question, students need to know that (1) double-stranded DNA is held together by weak hydrogen (H) bonding between members of a base pair; (2) in DNA replication, base pairs must separate to allow for copying; and (3) DNA replication prepares a cell for division by producing two complete copies of the genetic material. They also need the critical-thinking skill of **application** ***(applying their knowledge to determine why nitrogen mustard interferes with cell division and tumor growth). Complete answers will include each of the following (I, IIa, IIb, III, and IV): (I) An anticancer drug must somehow stop cell division, thus halting tumor growth. (IIa) Normally, bonding between DNA base pairs is weak (H bonding). (IIb) When DNA replication occurs, these weak bonds are broken (the DNA is "unzipped") to allow for copying, which is necessary for cell division. (III) Strong, covalent bonds between base pairs (as formed by nitrogen mustard) cannot be broken, so no DNA copying occurs and cell division ceases. (IV) Therefore, nitrogen mustard is an anticancer drug.***

(Fink 2003, Tuckman 2003, Juwah et al. 2004). Many students have communicated that they never understood how to "think critically" in other courses, even if they were asked to, because they were never shown what it means.

In addition to these benefits, we have found that student answers to these types of questions have provided exceptional formative feedback for us to refine and improve our assessments and teaching practices. In some cases, the feedback comes in the form of apparent student misconceptions of the course material. For instance, in example 1, we found that 36 percent of the students who opted for choice 2 either lacked a rationale or offered a rationale that was incorrect. These results suggested that many students were selecting an answer based on intuition or on a foggy memory of the material discussed in class, as opposed to sound analysis of the scenarios presented and application of the facts about membranes and RNA to those scenarios. To investigate this phenomenon further, we used a different form of the question later in the same semester, in which the molecule of interest was uncharged, which means that it can pass through the membrane without the use of a channel. Unfortunately, 43 percent of the class still proposed that this molecule would not pass through the membrane. Because this question was more complex, we expected that many students would have errors in reasoning. However, this large percentage was disappointing and may illustrate that many students failed to fully understand membrane structure, perhaps as a result of preconceived (and hard to change) notions about this material, or perhaps because of the manner in which they were exposed to it. Before we discovered this finding, the students learned about membranes through the standard combination of lectures and assigned readings in the textbook. Other components of the course are taught using alternative methods, such as peer interaction in lecture, role-playing, one-minute essays, and other techniques. We are changing the way we teach about membrane structure and function, and we will be using this same assessment methodology to measure the impact of these instructional changes on student understanding.

In other cases, the formative feedback arises when students demonstrate either limited understanding or unexpected insights about the material. In example 2, although we expected students to be most confused about the distinction between drift and selection, we found instead that the most common mistake was the failure to adequately describe genetic drift. Most students could only talk about genetic drift in terms of specific examples that create the conditions for genetic drift (e.g., founder effect). Many students even used the terms "genetic drift" and "founder effect" interchangeably. This type of revelation, which is a common result of the assessment methodology we are describing, allows for responsive changes in the grading rubric as well as in teaching approaches. For example, we might ask another class a version of the question in which the "I" population is moved onto the mainland, thus eliminating the easy rationale for assuming the founder effect.

In example 3, although the most predictable answer involves knowledge of DNA replication, an alternative approach begins with a different focus. The impact of nitrogen mustard is the same (it prevents unzipping of DNA), but the impact is on DNA transcription, not DNA replication. Since the cell cannot function without proper transcription, it is reasonable to assume that cell division would also cease, effectively stopping tumor growth. Here, the instructor can decide whether the question is broad enough to allow for this level of insight, especially if the student employs appropriate logic throughout. The rubric can be easily amended as desired for immediate or subsequent use. Alternatively, the question can be rewritten to further constrain the possible answers or to encourage an even greater diversity of responses. Since one characteristic of critical thinking is the awareness that a given question may have more than one correct answer, this methodology allows alternative answers to be considered and possibly built into the scoring rubric. Overall, this type of feedback has proved valuable in helping us identify specific areas of the course that need further refining to improve student learning and understanding.

Conclusions

We can imagine that some biology instructors might still be reluctant to use this methodology, despite its advantages for student learning, because of time constraints and other practical concerns. But our assessment methodology offers three particular advantages that can help alleviate these worries. First, these types of assessments demand content knowledge, so there are no "wasted" questions. Second, the assessments are flexible, in that they can be easily amended to accommodate unforeseen answers, and can be weighted to favor either the critical-thinking component or the content component. Third, the assessments can be more rapidly and reliably scored than other "open-ended" questions because of the highly refined format of the scoring rubrics.

We are currently studying individual gains in learning for both specific content (e.g., membrane structure and function, forces of evolution, and DNA replication) and critical-thinking skills (e.g., application or analysis) from one time point to another in a semester. Most instructors recognize that their discipline contains particular concepts that are known to be difficult for most students, and we are hoping that our investigations will clarify how to help students learn essential biological concepts more deeply.

We are also examining the transferability of skills developed in one context (e.g., introductory biology) to a different context (e.g., introductory physics). These types of investigations have created collaborative opportunities with instructors in other STEM disciplines at Duke University who are interested in improving student learning and curriculum goals. The critical-thinking assessments described here offer an entryway into understanding relationships between teaching practices, student learning, and assessment designs. We are currently parlaying this methodology into an interdisciplinary effort to enhance critical-thinking instruction across the

STEM disciplines at Duke University, guided in part by the collaborative working-group model described by Middendorf and Pace (2004) and Cox (2004). Although we are also working with other interested faculty within biology, we have found that these types of assessments are most needed and desired in the large introductory classes across disciplines, and conversations across disciplinary lines have helped faculty to see the value of assessing critical-thinking skills as a distinct goal from measuring content acquisition.

Acknowledgments

We would like to acknowledge the members of the Duke Consortium for the Scholarship of Teaching and Learning for their advice and feedback. In addition, we thank Jeremy Hyman for his contributions to the design of the questions used as examples. This work was supported in part by a grant to P. P. L. from Duke University's Arts and Sciences Committee on Faculty Research.

References cited

[AAAS] American Association for the Advancement of Science. 1989. Science for All Americans. New York: Oxford University Press.

Anderson LW, Krathwohl D, eds. 2001. A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. New York: Longman.

Aretz AJ, Bolen MT, Devereux KE. 1997. Critical thinking assessment of college students. Journal of College Reading and Learning 28: 12–23.

Aviles CB. 1999. Understanding and Testing for "Critical Thinking" with Bloom's Taxonomy of Educational Objectives. Washington (DC): ERIC Clearinghouse. Report no. SO032219.

Beyer BK. 1984. Improving thinking skills—defining the problem. Phi Delta Kappan 65: 486–490.

Bloom BS, ed. 1956. Taxonomy of Educational Objectives: The Classification of Educational Goals. New York: McKay.

Cox M. 2004. Introduction to faculty learning communities. New Directions for Teaching and Learning 97: 5–23.

Cromwell LS. 1992. Assessing critical thinking. New Directions for Community Colleges 77: 37–50.

Ebert-May D, Batzli J, Lim H. 2003. Disciplinary research strategies for assessment of learning. BioScience 53: 1221–1228.

Facione PA. 1990. Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction. Pages 1–19 in Facione PA, ed. The Delphi Report. Millbrae (CA): California Academic Press.

Facione PA, Facione NC, Giancarlo CA. 2000. The disposition toward critical thinking: Its character, measurement, and relationship to critical thinking skill. Informal Logic 20: 61–84.

Fink LD. 2003. Creating Significant Learning Experiences. San Francisco: Jossey-Bass.

Fisher A, Scriven M. 1997. Critical Thinking: Its Definition and Assessment. Point Reyes (CA): Edgepress.

Fox MA, Hackerman N, eds. 2003. Evaluating and Improving Undergraduate Teaching in Science, Technology, Engineering, and Mathematics. Washington (DC): National Academies Press.

Haas PF, Keeley SM. 1998. Coping with faculty resistance to teaching critical thinking. College Teaching 46: 63–67.

Handelsman J, et al. 2004. Scientific teaching. Science 304: 521–522.

Juwah C, Macfarlane-Dick D, Matthew B, Nicol D, Ross D, Smith B. 2004. Enhancing Student Learning through Effective Formative Feedback. York (United Kingdom): Higher Education Academy Generic Centre.

Middendorf J, Pace D. 2004. Decoding the disciplines: A model for helping students learn disciplinary ways of thinking. New Directions for Teaching and Learning 98: 1–12.

[NAS–NRC] National Academy of Sciences–National Research Council. 2003. BIO 2010: Transforming Undergraduate Education for Future Research Biologists. Washington (DC): National Academies Press.

Onwuegbuzie AJ. 2001. Critical thinking skills: A comparison of doctoral- and master's-level students. College Student Journal 35: 477–481.

Paul RW, Elder L, Bartell T. 1997. California Teacher Preparation for Instruction in Critical Thinking: Research Findings and Policy Recommendations. Santa Rosa (CA): Foundation for Critical Thinking.

Porter AC. 2002. Measuring the content of instruction: Uses in research and practice. Educational Researcher 31: 3–14.

Potts B. 1994. Strategies for Teaching Critical Thinking in Education. Washington (DC): ERIC Clearinghouse. Report no. ED385606.

Resnick LB. 1987. Education and Learning to Think. Washington (DC): National Academy Press.

Tuckman BW. 2003. The effect of learning and motivation strategies training on college students' achievement. Journal of College Student Development 44: 430–437.

Yuretich RF. 2004. Encouraging critical thinking: Measuring skills in large introductory science classes. Journal of College Science Teaching 33: 40–46.